lecture six

Yi (Chris) Chen

October 13, 2017

### lecture six

#### cross-validation

determine which model you should use when have no test data set: model selection

* some potential problems:

1. overfitting: too closely to specific properities of the training data, rather than the underlying distribution.  
   In KNN when k tends to be smaller, the error in training data tends to be smaller. (when k=1, the error in training data is 0 ) But this may tends to be larger error in testing data.

when you